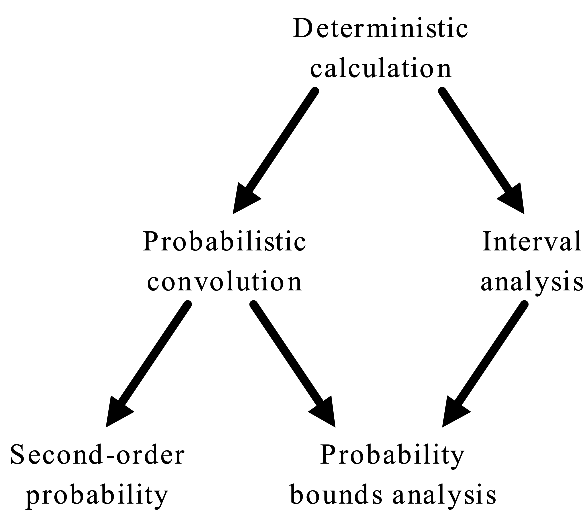
Probability bounds analysis is a global sensitivity analysis

Scott Ferson*, Janos Hajagos and W. Troy Tucker***

*Applied Biomathematics, 100 North Country Road, Setauket, New York 11733 USA, scott@ramas.com, 631-751-4350, fax -3435

**Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794 USA

Leamer (1990) defined global sensitivity analysis as a systematic study in which "a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified". There are two disparate ways to effect such a study. One natural way is to bound the neighborhood with interval ranges. Another natural way is to ascribe a probability distribution to the elements in the neighborhood. Consider, for example, the context of a deterministic calculation. When the model involves uncertainty about the real-valued quantities used in the calculation, the definition of global sensitivity analysis is equivalent to that of interval analysis (Dwyer 1951; Moore 1966; Alefield and Herzberger 1983; Neumaier 1990). Probability theory, implemented perhaps by Monte Carlo simulation, can also be viewed as a global sensitivity analysis of a deterministic calculation in that it yields a distribution describing the probability of alternative possible values about a point estimate (Iman and Helton 1985; Morgan and Henrion 1990; Helton and Davis 2002). In the figure below these two possible paths are shown as right and left downward arrows respectively.



Of course, the calculations on which it might be desirable to conduct sensitivity analyses are not all deterministic. In fact, many of them are already probabilistic, as is the case in most modern risk analyses and safety assessments. One could construct a probabilistic sensitivity analysis of a probabilistic calculation. The resulting analysis would be a second-order probabilistic assessment. However, such studies are often difficult to conduct because of the large number of calculations that are required. It is also sometimes difficult to visualize the results in a way that is easily comprehensible. Alternatively, one could apply bounding arguments to the probabilistic calculation and arrive at interval versions of probability distributions. We call such calculations "probability bounds analysis" (PBA; Ferson 1994; 2002; Ferson et al. 2003). This approach represents the uncertainty about a probability distribution by the set of cumulative distribution functions lying entirely within a pair of bounding distributions called a "probability box" or a "p-box". Probability bounds analysis is a global sensitivity analysis of a probabilistic calculation because it defines neighborhoods of probability distributions (i.e., the p-boxes) that represent the uncertainty about imperfectly known input distributions and projects this uncertainty through the model to identify a neighborhood of answers (also characterized by a p-box) in a way that guarantees the resulting bounds will entirely enclose the cumulative distribution function of the output. A probability distribution is to a p-box the same way a real scalar number is to an interval. The bounding distributions of the p-box enclose all possible distributions in the same way that the endpoints of the interval circumscribe the possible real values.

Probability bounds analysis is related to other forms of uncertainty analysis. It is a marriage of probability theory and interval analysis that generalizes and is faithful to both traditions. As depicted in the figure, PBA can arise either by bounding probability distributions (the left path down to PBA) or by forming probability distributions of intervals (the right path). The advantage of this marriage is that variability (aleatory uncertainty) and incertitude (epistemic uncertainty) are treated separately and propagated differently so each maintains its own character. PBA is a comprehensive global sensitivity analysis that is an alternative to complicated second-order or nested Monte Carlo methods. PBA is very similar in spirit to Bayesian sensitivity analysis (which is also known as robust Bayes; Insua and Ruggeri 2000), although the former concerns arithmetic and convolutions, and the latter addresses the issues of updating and aggregation. Unlike Bayesian sensitivity analysis, probability bounds analysis is always easy to employ because it does not depend on the use of conjugate pairs to make calculations simple. PBA is a simplified approach to computing with imprecise probabilities (Walley 1991). Like a Bayesian sensitivity analysis, imprecise probabilities are represented by a class of distribution functions. PBA is simpler because it defines the class solely by reference to two bounding distributions. (It therefore cannot represent a situation in which there are intermediate distributions lying within the bounds that are excluded from the class.) In the context of risk and safety assessments, however, this is rarely a significant drawback.

PBA can produce rigorous bounds around the output distribution from an assessment. These bounds enclose all the possible distributions that could actually arise given what is known and what is not known about the model and its inputs. Because it is based on the idea of bounding rather than approximation, it provides an estimate of its own reliability (Berleant 1993; 1996; cf. Adams and Kulisch, 1993). Probability bounds analysis can comprehensively account for possible deviations in assessment results arising from uncertainty about

- distribution parameters,
- distribution shape or family,
- intervariable dependence, and even
- model structure.

Moreover, it can handle all of these kinds of uncertainties in a single calculation that gives a simple and rigorous characterization of how different the result could be given all of the professed uncertainty. The requisite computations used in PBA are actually quite simple and have been implemented in straightforward algorithms (Yager 1986; Williamson and Downs 1990; Berleant 1993, 1996; Ferson et al. 2002; Ferson and Hajagos 2004). The computations are generally much faster than even simple Monte Carlo convolution and vastly faster than a numerically intensive sensitivity analysis with traditional methods (Iman and Conover 1980; 1982; Iman et al. 1981a; 1981b; Iman and Helton 1985; Morgan and Henrion 1990; Saltelli et al. 2001; Helton and Davis 2002). The full paper will outline, via pseudocode, how these algorithms work.

Probability bounds analysis is useful whenever the uncertainty about the marginal distributions can be characterized by interval bounds about the cumulative distribution function. These bounds can be specified using empirical information available about each distribution. For instance, if the parameters of a normal distribution can be given within interval ranges, best-possible bounds on the distribution are easy to construct. If the shape of the underlying distribution is not known, but some statistics such as the mean, mode, variance, etc. can be specified (or given as intervals), rigorous bounds can generally be constructed that are guaranteed to enclose the true distribution subject to the given constraints. Often these bounds will be optimally narrow given the stated information. The resulting p-boxes are distribution-free in the sense that they make no assumptions whatever about the distribution family (whether it is normal, lognormal, Weibull, etc.). Such bounds on distributions can then be combined according to the calculations in the assessment. Currently, software is available to handle

- arithmetic convolutions (addition, multiplication, minimum, etc.),
- magnitude comparisons (greater than, less than),
- logical operations (conjunction, disjunction, etc.), and
- transformations (logarithm, exponentiation, roots, etc.).

Unlike approaches based on conventional Monte Carlo simulation, the algorithms employed for these operations yield rigorous answers that lack sampling error. In fact, the results are exact at each point of discretization, of which there may be arbitrarily many. The results are guaranteed to enclose the true distributions.

It is also possible to handle uncertainty about the dependencies among variables in a model. Recent algorithmic developments permit uncertainty about the dependencies among variables to be propagated through the calculations of a probabilistic assessment (Ferson et al. 2004). Each pair-wise dependency may be modeled with any of the following six assumptions:

- independence,
- perfect (positive or negative) dependency,
- linear relationship and correlation within a specified interval,
- linear relationship with unknown correlation,
- signed (positive or negative) but otherwise unknown dependency, and
- unknown dependency (including any nonlinear relationship).

For the first two cases, a convolution between two probability distributions yields a well defined probability distribution. For the latter four cases, the results are given as bounds on a (cumulative) distribution function. The bounds obtained are generally the best possible bounds, i.e., they could not be any narrower yet still contain all the possible distributions permissible under the assumption.

It is also possible, of course, to perform a sensitivity analysis on the results of an assessment conducted with probability bounds analysis. The full paper will address how existing methods for sensitivity analysis can be generalized for use in probability bounds analysis and how a meta-level sensitivity analyses can be applied to probability bounds analysis.

Acknowledgments

This work was supported by funding from the National Institutes of Health and Sandia National Laboratories.

References

Alefeld, G. and J. Herzberger. 1983. *Introduction to Interval Computations.* Academic Press, New York.

Berleant, D. 1993. Automatically verified reasoning with both intervals and probability density functions. *Interval Computations* (1993 no. 2): 48-70.

Berleant, D. 1996. Automatically verified arithmetic on probability distributions and intervals. Pp. 227-244 in *Applications of Interval Computations*, edited by B. Kearfott and V. Kreinovich, Kluwer Academic Publishers.

Berleant, D. and B. Liu. 1997. Is sensitivity analysis more fault tolerant than point prediction? Pages 196-199 in *Simulation in the Medical Sciences*. Proceedings of the 1997 Western Multiconference, The Society for Computer Simulation International.

Dwyer, P. 1951. *Linear Computations*. John Wiley, New York.

Ferson, S. 2002. *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain Numbers.* Lewis Publishers, Boca Raton, Florida.

Ferson, S. and J. Hajagos. 2004. Arithmetic with uncertain numbers: rigorous and often best-possible answers. *Reliability and Engineering and System Safety* [to appear].

Ferson, S., V. Kreinovich, L. Ginzburg, K. Sentz and D.S. Myers. 2003. *Constructing Probability Boxes and Dempster-Shafer Structures*. Sandia National Laboratories, SAND2002-4015, Albuquerque, New Mexico.

Ferson, S., W.T. Tucker, D. Berleant, J. Zhang and L. Ginzburg. 2004 [tentative]. *Dependence in Dempster-Shafer Theory and Probability Bounds Analysis*. Sandia National Laboratories, SAND2004-XXXX, Albuquerque, New Mexico.

Frank, M.J., R.B. Nelsen, and B. Schweizer. 1987. Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* 74: 199-211.

Iman, R.L. and W.J. Conover. 1980. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics A9*:1749-1842.

Iman, R.L. and W.J. Conover. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics B11*:311-334.

Iman, R.L. and J.M. Davenport. 1982. *An Interactive Algorithm to Produce a Positive-definite Correlation Matrix from an Approximate Correlation Matrix (With a Program Users' Guide)*. SAND81-1376, Sandia National Laboratories, Albuquerque.

Iman, R.L. and J.C. Helton. 1985. *A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models*. NUREG/CR-3904. National Technical Information Service, Springfield, Virginia.

Iman, R.L and M.J. Shortencarier. 1984. *A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models*. Technical Report NUREG/CR-3624, SAND83-2365, Sandia National Laboratories, Albuquerque, New Mexico.

Iman, R.L., J.C. Helton, J.E. Campbell. 1981a. An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assessment. *Journal of Quality Technology* 13:174-183.

Iman, R.L., J.C. Helton, J.E. Campbell. 1981b. An approach to sensitivity analysis of computer models, Part 2. Ranking of input variables, response surface validation, distribution effect and technique synopsis. *Journal of Quality Technology* 13:232-240.

Insua, D.R. and F. Ruggeri (eds.) 2000. *Robust Bayesian Analysis*. Lecture Notes in Statistics, Volume 152. Springer-Verlag, New York.

Leamer, E.E. 1990. Sensitivity analysis would help. *Modelling Economic Series*, edited by C.W.J. Granger, Clarendon Press, Oxford.

Moore, R.E. 1966. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Morgan, M.G. and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, England.

Neumaier, A. 1990. *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge.

Saltelli, A. 2002. Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145: 280-297.

Saltelli, A. and S. Tarantola. 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association* 97(459): 702-709.

Saltelli, A., S. Tarantola and F. Campolongo. 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15(4): 377-395.

Saltelli, A., K. Chan, E.M. Scott (eds.) 2001. *Sensitivity Analysis*. John Wiley, New York.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Yager, R.R. 1986. Arithmetic and other operations on Dempster-Shafer structures. *International Journal of Man-machine Studies* 25: 357-366.